

ON THE  $k$ -ABELIAN COMPLEXITY OF THE CANTOR SEQUENCE

JIN CHEN, XIAOTAO LÜ, AND WEN WU\*

ABSTRACT. In this paper, we prove that for every integer  $k \geq 1$ , the  $k$ -abelian complexity function of the Cantor sequence  $\mathbf{c} = 101000101 \dots$  is a 3-regular sequence.

## 1. INTRODUCTION

This paper is devoted to the study of the  $k$ -abelian complexity of the Cantor sequence

$$\mathbf{c} := c_0 c_1 c_2 \dots = 101000101000000000101000101 \dots$$

which satisfies  $c_0 = 1$  and for all  $n \geq 0$ ,

$$c_{3n} = c_{3n+2} = c_n \text{ and } c_{3n+1} = 0. \quad (1.1)$$

The  $k$ -abelian complexity, which was introduced by Karhumäki in [8], is a measure of disorder of infinite words. It has been studied widely in [12, 13, 14, 15, 16]. Before we give its definition, we need some notations. Let  $\mathcal{A}$  be a finite alphabet and  $\mathcal{A}^n$  be the set of words of length  $n$  for every positive integer  $n$ . Denote  $\mathcal{A}^*$  the set of all finite words on  $\mathcal{A}$ . For two words  $u, v \in \mathcal{A}^*$ ,  $v$  is called a *factor* of  $u$  if  $u = wvw'$  where  $w, w' \in \mathcal{A}^*$ . For a word  $u = u_0 u_1 \dots u_{n-1} \in \mathcal{A}^n$ , the *prefix* and *suffix* of length  $\ell \geq 1$  are defined as

$$\text{pref}_\ell(u) := u_0 u_1 \dots u_{\ell-1} \text{ and } \text{suff}_\ell(u) := u_{n-\ell} \dots u_{n-1};$$

while for  $\ell \leq 0$ , we define  $\text{pref}_\ell(u) = \varepsilon$  and  $\text{suff}_\ell(u) = \varepsilon$ , where  $\varepsilon$  is the empty word. Denote  $|u|$  the length of a word  $u$  and denote  $|u|_v$  the number of occurrences of a word  $v$  in  $u$ .

**Definition 1** (see [17]). Let  $k \geq 1$  be an integer. Two words  $u, v \in \mathcal{A}^*$  are called  *$k$ -abelian equivalent*, written by  $u \sim_k v$ , if  $\text{pref}_{k-1}(u) = \text{pref}_{k-1}(v)$ ,  $\text{suff}_{k-1}(u) = \text{suff}_{k-1}(v)$  and  $|u|_w = |v|_w$  for every  $w \in \mathcal{A}^k$ .

The above definition is one of the equivalent definitions of the  $k$ -abelian equivalence; see also [16]. The  $k$ -abelian equivalence is in fact an equivalence relation. The  *$k$ -abelian complexity* of an infinite word  $\omega$  is the function  $\mathcal{P}_\omega^{(k)} : \mathbb{N} \rightarrow \mathbb{N}$  and for every  $n \geq 1$ ,  $\mathcal{P}_\omega^{(k)}(n)$  is assigned to be the number of  $k$ -abelian equivalence classes of factors of  $\omega$  of length  $n$ . Precisely, for every positive integer  $n$ ,

$$\mathcal{P}_\omega^{(k)}(n) = \text{Card}(\mathcal{F}_\omega(n) / \sim_k),$$

where  $\mathcal{F}_\omega(n)$  is the set of all factors of length  $n$  occurring in  $\omega$ .

In our first result, we reduce the  $k$ -abelian equivalence of any two factors of  $\mathbf{c}$  to the abelian equivalence of such factors. In detail, we prove the following theorem.

**Theorem 1.** *Let  $k \geq 1$  be an integer and let  $u, v$  be two factors of  $\mathbf{c}$  satisfying  $|u| = |v|$ . If  $\text{pref}_k(u) = \text{pref}_k(v)$  and  $\text{suff}_k(u) = \text{suff}_k(v)$ , then  $u \sim_{k+1} v$  if and only if  $u \sim_1 v$ .*

2010 *Mathematics Subject Classification.* 68R15 and 11B85.

*Key words and phrases.* Cantor sequence;  $k$ -abelian complexity;  $k$ -regular sequences.

This work was supported by NSFC (Nos. 11401188 and 11626110).

\* Wen Wu is the corresponding author.

By using Theorem 1, we are able to study the  $k$ -abelian complexity of  $\mathbf{c}$  for every  $k \geq 1$ , and we have the following result.

**Theorem 2.** *For every integer  $k \geq 1$ , the  $k$ -abelian complexity function of the Cantor sequence is a 3-regular sequence.*

The  $k$ -regular sequence was introduced by Allouch and Shallit [2] as an extension of the  $k$ -automatic sequence. The definitions of the  $k$ -automatic sequences and the  $k$ -regular sequences are stated below; see also [1, 6].

**Definition 2.** For an integer  $k \geq 1$ , a sequence  $\mathbf{w} = (w_n)_{n \geq 0}$  is a  $k$ -automatic sequence if its  $k$ -kernel

$$\mathcal{K}_k(\mathbf{w}) = \{(w_{k^e n + c})_{n \geq 0} \mid e \geq 0, 0 \leq c < k^e\}$$

is a finite set. The sequence  $\mathbf{w}$  is called a  $k$ -regular sequence if the  $\mathbb{Z}$ -module generated by its  $k$ -kernel is finitely generated.

Karhumäki, Saarela and Zamboni [14] studied the  $k$ -abelian complexity of the Thue-Morse sequence, which is a 2-automatic sequence. Vandomme, Parreau and Rigo [17] conjectured that the 2-abelian complexity of the Thue-Morse sequence is a 2-regular sequence. This has been proved independently in [12] by Greinecker and in [13] by Parreau, Rigo, Rowland and Vandomme.

Our result (Theorem 2) supports the following more general conjecture, which has been posed in [13].

**Conjecture 1.** *The  $k$ -abelian complexity of any  $\ell$ -automatic sequence is an  $\ell$ -regular sequence.*

This paper is organized as follows. In Section 2, we give the recurrence relations for the abelian complexity function of the sequence  $\mathbf{c}$ . As a consequence, the abelian complexity function of the Cantor sequence is a 3-regular sequence. In Section 3, we prove Theorem 1. In the last section, we give the proof of Theorem 2.

## 2. ABELIAN COMPLEXITY

The abelian complexity of an infinite word  $\omega$  is in fact the 1-abelian complexity of  $\omega$ . For more details of the abelian complexity, see [4, 5, 8, 9, 10, 11] and references therein. In this section, we shall investigate the abelian complexity of  $\mathbf{c}$ .

First we introduce a useful result which characterizes the left and right special factors of  $\mathbf{c}$ . Recall that a factor  $v$  of  $w$  is called *right special* (resp. *left special*) if both  $va$  and  $vb$  (resp.  $av$  and  $bv$ ) are factors of  $w$  for distinct letters  $a, b \in \mathcal{A}$ . We denote  $\mathcal{RS}_w(n)$  (resp.  $\mathcal{LS}_w(n)$ ) the set of all right special (resp. left special) factors of  $w$  of length  $n$ .

**Lemma 1.** *For every  $i \geq 0$  and  $3^i < k \leq 3^{i+1}$ ,*

$$\mathcal{RS}_{\mathbf{c}}(k) = \{0^k, \text{suff}_k(\sigma^i(010))\} \text{ and } \mathcal{LS}_{\mathbf{c}}(k) = \{0^k, \text{pref}_k(\sigma^i(010))\}.$$

*Proof.* The result follows from [7, Theorem 1] and the fact that every left special factor in  $\mathbf{c}$  is the reversal of some right special factor in  $\mathbf{c}$ .  $\square$

Let  $\omega = \omega_0\omega_1\omega_2\cdots$  be an infinite sequence on  $\{0, 1\}$ . It is proved in [3, Proposition 2.2] that the abelian complexity of  $\omega$  is related to its digit sums in the following way: for every  $n \geq 1$ ,

$$\mathcal{P}_{\omega}^{(1)}(n) = M_{\omega}(n) - m_{\omega}(n) + 1, \tag{2.1}$$

where

$$M_{\omega}(n) := \max \{ \sum_{j=i}^{i+n-1} \omega_j \mid i \geq 0 \} \text{ and } m_{\omega}(n) := \min \{ \sum_{j=i}^{i+n-1} \omega_j \mid i \geq 0 \}.$$

For the digit sums of the Cantor sequence  $\mathbf{c}$ , we have the following lemma.

**Lemma 2.** For every integer  $n \geq 1$ ,  $M_{\mathbf{c}}(n) = \sum_{i=0}^{n-1} c_i$  and  $m_{\mathbf{c}}(n) = 0$ .

*Proof.* Since  $0^n$  is always a factor of  $\mathbf{c}$  for every  $n \geq 1$ , we have  $m_{\mathbf{c}}(n) = 0$  for every  $n \geq 1$ .

For every  $i \geq 0$  and  $n \geq 1$ , let  $\Sigma(i, n) := \sum_{j=i}^{i+n-1} c_j$ . We only need to show that  $M_{\mathbf{c}}(n) \leq \Sigma(0, n)$  for every  $n \geq 1$ , since the inverse inequality always holds by definition. For this purpose, we shall prove that for every  $n \geq 1$ ,

$$\Sigma(i, n) \leq \Sigma(0, n) \text{ for every integer } i \geq 0. \quad (2.2)$$

Since '1' occurs in  $\mathbf{c}$  and '11' does not occur in  $\mathbf{c}$ , we have  $\Sigma(i, 1) \leq 1 = \Sigma(0, 1)$  and  $\Sigma(i, 2) \leq 1 = \Sigma(0, 2)$ . Now suppose (2.2) holds for  $n < m$ . We first deal with the case:  $m = 3j + 2$ . By (1.1), we have the following nine recurrence relations:

$$\begin{cases} \Sigma(3i, 3n) = 2\Sigma(i, n), & \Sigma(3i+1, 3n+2) = \Sigma(i, n+1) + \Sigma(i+1, n), \\ \Sigma(3i, 3n+1) = \Sigma(i, n) + \Sigma(i, n+1), & \Sigma(3i+2, 3n) = \Sigma(i, n) + \Sigma(i+1, n), \\ \Sigma(3i, 3n+2) = \Sigma(i, n) + \Sigma(i, n+1), & \Sigma(3i+2, 3n+1) = \Sigma(i, n+1) + \Sigma(i+1, n), \\ \Sigma(3i+1, 3n) = \Sigma(i, n) + \Sigma(i+1, n), & \Sigma(3i+2, 3n+2) = \Sigma(i, n+1) + \Sigma(i+1, n+1), \\ \Sigma(3i+1, 3n+1) = \Sigma(i, n) + \Sigma(i+1, n). \end{cases}$$

By the above equations and the inductive assumption, for every  $i \geq 0$ ,

$$\begin{aligned} \Sigma(3i, 3j+2) &= \Sigma(i, j) + \Sigma(i, j+1) \leq \Sigma(0, j) + \Sigma(0, j+1) = \Sigma(0, 3j+2), \\ \Sigma(3i+1, 3j+2) &= \Sigma(i+1, j) + \Sigma(i, j+1) \leq \Sigma(0, j) + \Sigma(0, j+1) = \Sigma(0, 3j+2). \end{aligned}$$

Note that at least one of  $c_i$  and  $c_{i+1}$  must be zero. So

$$\Sigma(3i+2, 3j+2) = \Sigma(i, j+1) + \Sigma(i+1, j+1) \leq \Sigma(0, j) + \Sigma(0, j+1) = \Sigma(0, 3j+2).$$

Therefore, (2.2) holds in the case  $m = 3j + 2$ . Following the same way, we can verify (2.2) when  $m = 3j, 3j + 1$ .  $\square$

**Corollary 1.**  $M_{\mathbf{c}}(1) = 1$ ,  $M_{\mathbf{c}}(2) = 1$  and for every  $n \geq 1$ ,

$$M_{\mathbf{c}}(3n) = 2M_{\mathbf{c}}(n) \text{ and } M_{\mathbf{c}}(3n+1) = M_{\mathbf{c}}(3n+2) = M_{\mathbf{c}}(n) + M_{\mathbf{c}}(n+1).$$

Moreover,  $\{M_{\mathbf{c}}(n)\}_{n \geq 1}$  is a 3-regular sequence.

**Proposition 1.**  $\mathcal{P}_{\mathbf{c}}^{(1)}(1) = 2$ ,  $\mathcal{P}_{\mathbf{c}}^{(1)}(2) = 2$  and for every  $n \geq 1$ ,

$$\mathcal{P}_{\mathbf{c}}^{(1)}(3n) = 2\mathcal{P}_{\mathbf{c}}^{(1)}(n) - 1 \text{ and } \mathcal{P}_{\mathbf{c}}^{(1)}(3n+1) = \mathcal{P}_{\mathbf{c}}^{(1)}(3n+2) = \mathcal{P}_{\mathbf{c}}^{(1)}(n) + \mathcal{P}_{\mathbf{c}}^{(1)}(n+1) - 1.$$

Moreover,  $\{\mathcal{P}_{\mathbf{c}}^{(1)}(n)\}_{n \geq 1}$  is a 3-regular sequence.

*Proof.* It follows from Lemma 2, Corollary 1 and (2.1).  $\square$

### 3. FROM $k$ -ABELIAN EQUIVALENCE TO 1-ABELIAN EQUIVALENCE

In this section, we give a key theorem, which implies that under certain condition,  $k$ -abelian equivalence can be reduced to 1-abelian equivalence. Using this theorem, we deduce the regularity of the  $k$ -abelian complexity of  $\mathbf{c}$  from that of the abelian complexity of  $\mathbf{c}$ . Before stating the result, we give two auxiliary lemmas. For  $z, w \in \mathcal{A}^*$ , we define

$$P(z, w) := \begin{cases} 1, & \text{if } z \text{ is a prefix of } w, \\ 0, & \text{otherwise,} \end{cases} \text{ and } S(z, w) := \begin{cases} 1, & \text{if } z \text{ is a suffix of } w, \\ 0, & \text{otherwise.} \end{cases}$$

**Lemma 3.** Let  $\omega \in \{0, 1\}^{\mathbb{N}}$  and  $u, z \in \mathcal{F}_\omega$  with  $|u| \geq |z|$ . Suppose  $z = ayb$ , where  $a, b \in \{0, 1\}$ . We have

$$|u|_z = \begin{cases} |u|_{ay} - S(ay, u), & \text{if } ay \notin RS_\omega, \\ |u|_{yb} - P(yb, u), & \text{if } yb \notin LS_\omega, \\ |u|_{ay} - |u|_{ay(1-b)} - S(ay, u), & \text{if } ay \in RS_\omega, \\ |u|_{yb} - |u|_{(1-a)yb} - P(yb, u), & \text{if } yb \in LS_\omega. \end{cases}$$

*Proof.* Note that  $|u|_{ay} - S(ay, u)$  is the number of occurrences of a right extendable  $ay$  in  $u$ . When  $ay$  is not right special, every right extension of a right extendable  $ay$  must be  $z$ . So,  $|u|_{ay} - S(ay, u) = |u|_z$ . When  $ay$  is right special, its right extensions are either  $z$  or  $ay(1-b)$ . So,  $|u|_{ay} - S(ay, u) = |u|_z + |u|_{ay(1-b)}$ . The rest cases can be verified in the same way.  $\square$

**Lemma 4.** For every  $i \geq 0$ ,  $u \in \mathcal{F}_c$ , let  $\Delta_i := |u|_{0^{3^i+2}} + |u|_{10^{3^i}1} - |u|_{0^{3^i+1}} + 1$ . Then  $\Delta_i \in \{0, 1, 2\}$  and

$$\Delta_i = \begin{cases} |u|_{0^{3^i}1} + \frac{2}{3^i}|u|_{0^{3^i+1}} + 1 - P(0^{3^i}1, u) \pmod{3}, & \text{if } P(0^{3^i+1}, u) = S(0^{3^i+1}, u) = 0, \\ |u|_{0^{3^i}1} + 1 - S(0^{3^i+1}, u) - P(0^{3^i}1, u) \pmod{2}, & \text{otherwise.} \end{cases}$$

*Proof.* Let  $Z(\ell)$  ( $\ell \geq 1$ ) be the number of blocks of zeros (in  $u$ ) of length not less than  $\ell$ . For example, when  $u = 0010100$ , then  $Z(1) = 3$  and  $Z(2) = 2$ . Note that, for every  $\ell \geq 3^i + 1$ ,  $|0^\ell|_{0^{3^i+1}} - |0^\ell|_{0^{3^i+2}} = 1$ . So,

$$\begin{aligned} |u|_{0^{3^i+1}} - |u|_{0^{3^i+2}} &= \sum_{v \text{ is a block of zeros in } u} (|v|_{0^{3^i+1}} - |v|_{0^{3^i+2}}) \\ &= \sum_{\substack{v \text{ is a block of zeros in } u \\ |v| \geq 3^i+1}} 1 = Z(3^i + 1). \end{aligned}$$

On the other hand,  $10^{3^i}1$  only occurs in  $\sigma^{i+1}(1)$ . Thus, there is a block of zeros of length  $3^{i+\ell}$  (for some  $\ell \geq 1$ ) between two consecutive  $10^{3^i}1$ . Since the block of zeros could also be the prefix or suffix of  $u$ , we have  $|u|_{10^{3^i}1} - 1 \leq Z(3^i + 1) \leq |u|_{10^{3^i}1} + 1$ , which implies  $\Delta_i \in \{0, 1, 2\}$ .

When  $P(0^{3^i+1}, u) = 1$  or  $S(0^{3^i+1}, u) = 1$ , there is at least one block of zeros of length not less than  $3^i + 1$ , which is not located between two consecutive  $10^{3^i}1$ . This implies that  $|u|_{10^{3^i}1} \leq Z(3^i + 1) \leq |u|_{10^{3^i}1} + 1$ . So, in this case,  $\Delta_i \in \{0, 1\}$ . Applying Lemma 3 to  $|u|_{0^{3^i+2}}$  and  $|u|_{10^{3^i}1}$ , we have

$$\Delta_i = |u|_{0^{3^i}1} - 2|u|_{0^{3^i+1}} + 1 - S(0^{3^i+1}, u) - P(0^{3^i}1, u). \quad (3.1)$$

Since  $\Delta_i \in \{0, 1\}$ , by (3.1),  $\Delta_i = |u|_{0^{3^i}1} + 1 - S(0^{3^i+1}, u) - P(0^{3^i}1, u) \pmod{2}$ .

Now, suppose  $P(0^{3^i+1}, u) = S(0^{3^i+1}, u) = 0$ . Applying Lemma 3 to  $|u|_{0^{3^i+1}}$ , by (3.1), we have

$$\Delta_i = |u|_{0^{3^i}1} - 2Z(3^i + 1) + 1 - P(0^{3^i}1, u). \quad (3.2)$$

Let  $\sum_v$  denote the sum over all blocks of zeros  $v$  of  $u$  of length not less than  $3^i + 1$ . Then

$$|u|_{0^{3^i+1}} = \sum_v |v|_{0^{3^i+1}} = \sum_v (|v| - 3^i) = \left( \sum_v |v| \right) - 3^i Z(3^i + 1)$$

Note that, in this case, all blocks of zeros of  $u$  are of length  $3^{i+\ell}$  for some  $\ell \geq 1$ . So,

$$-2Z(3^i + 1) \equiv \frac{2}{3^i} |u|_{0^{3^i+1}} \pmod{3}. \quad (3.3)$$

The result of this case follows from (3.3) and (3.2).  $\square$

Now, we prove Theorem 1.

*Proof of Theorem 1.* Let  $u, v \in \mathcal{F}_{\mathbf{c}}$  satisfying  $|u| = |v|$ ,  $\text{pref}_k(u) = \text{pref}_k(v)$  and  $\text{suff}_k(u) = \text{suff}_k(v)$ . When  $k \geq |u|$ , the assumption gives  $u = v$ . In this case, the result is trivial. In the following, we always assume that  $k < |u|$ .

The ‘only if’ part follows directly from the definition of  $k$ -abelian equivalence. For the ‘if’ part, we only need to show that  $u \sim_k v$  implies that for every  $z \in \mathcal{F}_{\mathbf{c}}(k+1)$ ,  $|u|_z = |v|_z$ . For this purpose, we separate  $\mathcal{F}_{\mathbf{c}}(k+1)$  into two disjoint parts, i.e.,  $\mathcal{F}_{\mathbf{c}}(k+1) = E_1 \cup E_2$ , where

$$\begin{aligned} E_1 &= \{z \in \mathcal{F}_{\mathbf{c}}(k+1) \mid \text{pref}_k(z) \notin \mathcal{RS}_{\mathbf{c}}(k) \text{ or } \text{suff}_k(z) \notin \mathcal{LS}_{\mathbf{c}}(k)\}, \\ E_2 &= \{z \in \mathcal{F}_{\mathbf{c}}(k+1) \mid \text{pref}_k(z) \in \mathcal{RS}_{\mathbf{c}}(k) \text{ and } \text{suff}_k(z) \in \mathcal{LS}_{\mathbf{c}}(k)\}. \end{aligned}$$

Suppose  $z \in E_1$ . If  $\text{pref}_k(z) \notin \mathcal{RS}_{\mathbf{c}}(k)$ , then by Lemma 3,

$$|u|_z = |u|_{\text{pref}_k(z)} - S(\text{pref}_k(z), u) = |v|_{\text{pref}_k(z)} - S(\text{pref}_k(z), v) = |v|_z.$$

If  $\text{suff}_k(z) \notin \mathcal{LS}_{\mathbf{c}}(k)$ , then by Lemma 3,

$$|u|_z = |u|_{\text{suff}_k(z)} - P(\text{suff}_k(z), u) = |v|_{\text{suff}_k(z)} - P(\text{suff}_k(z), v) = |v|_z.$$

So, for every  $z \in E_1$ ,  $|u|_z = |v|_z$ .

Now, let  $z \in E_2$ . Suppose  $3^i < k \leq 3^{i+1}$  for some  $i \geq 0$ . When  $k \neq 3^i + 1$ , by Lemma 1,  $E_2 = \{0^{k+1}\}$ . By Lemma 3 and the assumptions of this result,

$$\begin{aligned} |u|_{0^{k+1}} &= |u|_{0^k} - |u|_{0^{k-1}} - S(0^k, u) \\ &= |u|_{0^k} - (|u|_{0^{k-1}} - P(0^{k-1}, u)) - S(0^k, u) \\ &= |v|_{0^k} - (|v|_{0^{k-1}} - P(0^{k-1}, v)) - S(0^k, v) = |v|_{0^{k+1}}. \end{aligned}$$

When  $k = 3^i + 1$ , by Lemma 1,  $E_2 = \{0^{k+1}, 0^k 1, 10^k, 10^{k-1} 1\}$ . For every  $w \in \mathcal{F}_{\mathbf{c}}$ , by Lemma 3 and 4, we have the following linear system:

$$\begin{cases} |w|_{0^{k+1}} + |w|_{0^k 1} = |w|_{0^k} - S(0^k, w), \\ |w|_{0^{k+1}} + |w|_{10^k} = |w|_{0^k} - P(0^k, w), \\ |w|_{10^k} + |w|_{10^{k-1} 1} = |w|_{10^{k-1}} - S(10^{k-1}, w), \\ |w|_{0^{k+1}} + |w|_{10^{k-1} 1} = |w|_{0^k} - 1 + \Delta_i, \end{cases} \quad (3.4)$$

which determines  $(|w|_z)_{z \in E_2}$  uniquely. If  $u \sim_k v$ , then the linear systems (3.4) for  $u$  and  $v$  turn out to be the same one. So,  $u \sim_k v$  implies  $|u|_z = |v|_z$  for every factor  $z \in E_2$ .  $\square$

We may now apply Theorem 1 repeatedly to reduce the  $k$ -abelian equivalence to the 1-abelian equivalence under the condition of Theorem 1.

**Corollary 2.** *Let  $k \geq 1$  and  $u, v \in \mathcal{F}_{\mathbf{c}}$  satisfying  $|u| = |v|$ . If  $\text{pref}_k(u) = \text{pref}_k(v)$  and  $\text{suff}_k(u) = \text{suff}_k(v)$ , then  $u \sim_{k+1} v$  if and only if  $u \sim_1 v$ .*

**Remark 1.** A similar result for Sturmian words is obtained by Karhumäki, Saarela and Zamboni [16, Corollary 3.1]. We would like to ask that in general, what kind of infinite words share a property similar to Corollary 2?

#### 4. $k$ -ABELIAN COMPLEXITY

In this section, we first give the regularity of the 2-abelian complexity of  $\mathbf{c}$ . Then, by using Theorem 1 properly, we deduce the regularity of the  $k$ -abelian complexity of  $\mathbf{c}$ . We start by classifying the  $k$ -abelian equivalent classes of  $\mathcal{F}_{\mathbf{c}}(n)$  by their prefixes and suffixes of length  $k-1$ .

For every  $k \geq 2$ ,  $x, y \in \mathcal{F}_{\mathbf{c}}(k-1)$  and every  $n \geq 1$ , let

$$p_k(n, x, y) := \text{Card}(\mathcal{W}_{n,x,y} / \sim_k),$$

where

$$\mathcal{W}_{n,x,y} := \{w \in \mathcal{F}_{\mathbf{c}}(n) \mid \text{pref}_{k-1}(w) = x, \text{suff}_{k-1}(w) = y\}.$$

Here  $p_k(n, x, y)$  denotes the number of  $k$ -abelian equivalent classes with the prefix  $x$  and the suffix  $y$ . Then, for every  $n \geq 1$ ,

$$\mathcal{P}_{\mathbf{c}}^{(k)}(n) = \sum_{x,y \in \mathcal{F}_{\mathbf{c}}(k-1)} p_k(n, x, y). \quad (4.1)$$

By Theorem 1,

$$\begin{aligned} p_k(n, x, y) &= \text{Card}(\mathcal{W}_{n,x,y} / \sim_k) \\ &= \text{Card}(\mathcal{W}_{n,x,y} / \sim_1) = \text{Card}(\{|w|_1 \mid w \in \mathcal{W}_{n,x,y}\}). \end{aligned} \quad (4.2)$$

**4.1. Regularity of the 2-abelian complexity of  $\mathbf{c}$ .** Recall that the Cantor sequence  $\mathbf{c}$  is the fixed point of the morphism  $\sigma : 0 \mapsto 000, 1 \mapsto 101$  starting by 1, i.e.,  $\mathbf{c} = \sigma^\infty(1)$ .

**Lemma 5.** *For all  $i, j \geq 1$ , let  $d_j$  be the number of '0' between the  $j$ -th '1' and the  $(j+1)$ -th '1' in  $\mathbf{c}$ , and let  $f(i, j) = j + \sum_{\ell=i}^{i+j-1} d_\ell$ . Then, for every  $j \geq 1$ ,*

$$d_{2j-1} = 1 \text{ and } d_{2j} = 3d_j. \quad (4.3)$$

Moreover, for all  $i, j \geq 1$ ,

$$\begin{cases} f(2i, 2j) = 3f(i, j), & f(2i, 2j+1) = 3f(i, j+1) - 2, \\ f(2i+1, 2j) = 3f(i+1, j), & f(2i+1, 2j+1) = 3f(i+1, j) + 2. \end{cases} \quad (4.4)$$

*Proof.* While applying  $\sigma$  to '1' or a block of '0's, we obtain only one block of '0's in both cases. Note that in  $\mathbf{c}$ , every '1' is followed by a block of '0's. Before the  $i$ -th '1', the number of occurrences of '1' is  $(i-1)$  and there are  $(i-1)$  blocks of '0's in  $\mathbf{c}$ . So, while applying  $\sigma$  to  $\mathbf{c}$ , the  $i$ -th '1' will generate the  $(2i-1)$ -th block of '0's, which implies  $d_{2i-1} = 1$ . For the same reason, the  $i$ -th block of '0's will generate the  $2i$ -th block of '0's. So,  $d_{2i} = 3d_i$ . This proves (4.3).

The recurrence relations (4.4) follows directly from (4.3). We verify the first one as an example:

$$f(2i, 2j) = 2j + \sum_{\ell=2i}^{2i+2j-1} d_\ell = 2j + \sum_{\ell=i}^{i+j-1} (d_{2\ell} + d_{2\ell+1}) = 3j + 3 \sum_{\ell=i}^{i+j-1} d_\ell = 3f(i, j).$$

□

**Proposition 2.**  $p_2(1, 0, 0) = p_2(1, 1, 1) = 1$ ,  $p_2(1, 0, 1) = p_2(1, 1, 0) = 0$  and for every  $n \geq 2$ ,

$$\begin{cases} p_2(n, 0, 0) = M_{\mathbf{c}}(n-2) + 1, \end{cases} \quad (4.5a)$$

$$\begin{cases} p_2(n, 1, 0) = p_2(n, 0, 1) = M_{\mathbf{c}}(n-1), \end{cases} \quad (4.5b)$$

$$\begin{cases} p_2(n, 1, 1) = \begin{cases} 0, & \text{if } n \equiv 0 \pmod{2}, \\ 1, & \text{if } n \equiv 1 \pmod{2}. \end{cases} \end{cases} \quad (4.5c)$$

*Proof.* The initial values can be showed by enumerating all the factors of length 1 and 2. Now, let  $n \geq 2$  and suppose  $n < 3^i$  for some  $i \geq 1$ .

Clearly, for every  $w \in \mathcal{W}_{n,0,0}$ ,  $|w|_1 \leq M_{\mathbf{c}}(n-2)$ . So,  $p_2(n, 0, 0) \leq M_{\mathbf{c}}(n-2) + 1$ . We prove the inverse inequality in the following. For every  $0 \leq \ell \leq n-1$ , let  $W_\ell = 0^{n-\ell} \text{pref}_\ell(\sigma^i(1))$  that is a factor of  $\sigma^i(01)$  and hence, a factor of  $\mathbf{c}$ . Note that  $|W_0|_1 = 0$  and  $|W_{n-2}|_1 = M_{\mathbf{c}}(n-2)$ . Since  $|W_\ell|_1 \leq |W_{\ell+1}|_1 \leq |W_\ell|_1 + 1$ , we know that  $|W_\ell|_1$  changes continuously from 0 to  $M_{\mathbf{c}}(n-2)$  while  $\ell$  takes values from 0 to  $n-2$ . Therefore, for every  $0 \leq s \leq M_{\mathbf{c}}(n-2)$ , there exists  $0 \leq \ell \leq n-2$  such that  $|W_\ell|_1 = s$ . If the last letter of  $W_\ell$  is 0, then  $W_\ell \in \mathcal{W}_{n,0,0}$ . Otherwise,  $|W_{\ell+1}|_1 = |W_\ell|_1 = s$  since 11 is not a factor of  $\mathbf{c}$ . So,  $W_{\ell+1} \in \mathcal{W}_{n,0,0}$ . This implies that  $p_2(n, 0, 0) \geq M_{\mathbf{c}}(n-2) + 1$  which proves (4.5a).

Since for every factor  $w$  of  $\mathbf{c}$ , its reversal  $\bar{w}$  is also a factor of  $\mathbf{c}$ , we have  $p_2(n, 1, 0) = p_2(n, 0, 1)$ . Then, applying a similar argument on the words  $W'_\ell = \text{suff}_\ell(\sigma^i(1))0^{n-\ell}$  where  $1 \leq \ell \leq n-1$ , we obtain (4.5b).

(In the rest of the proof, the symbol ' $\equiv$ ', otherwise stated, means equality modulo 2.)

Now, we prove (4.5c) for the case  $n \equiv 0$ . We first observe that for every  $w \in \mathcal{W}_{n,1,1}$ ,  $|w| \equiv 1$ . Since the number of 0 between two successive 1 must be  $3^j$  for some  $j \geq 0$  and  $3^j \equiv 1$ , we have  $|w|_0 \equiv |w|_1 - 1$  for every  $w \in \mathcal{W}_{n,1,1}$ . Therefore,  $|w| = |w|_0 + |w|_1 \equiv 1$ . Hence,  $\mathcal{W}_{n,1,1} = \emptyset$  when  $n$  is an even number, which implies  $p_2(n, 1, 1) = 0$  when  $n \equiv 0$ .

In the following, we will prove (4.5c) when  $n \equiv 1$ . For every  $w \in \mathcal{W}_{n,1,1}$ ,

$$n = |w| = |w|_1 + |w|_0 = 1 + f(i, |w|_1 - 1)$$

for some  $i \geq 1$ . (Since if a word occurs in  $\mathbf{c}$ , then it will occur infinitely many times in  $\mathbf{c}$ , we can assume  $i \geq 3$ .) Therefore, we only need to prove that for every  $m \geq 1$ , there is only one integer  $t_m \geq 2$  satisfying

$$2m + 1 = 1 + f(i, t_m) \quad (4.6)$$

for some  $i \geq 1$ . We reason by induction. Since  $\mathcal{W}_{3,1,1} = \{101\}$  and  $\mathcal{W}_{5,1,1} = \{10001\}$ , it follows that (4.6) holds for  $m = 1$  and 2. Assuming that (4.6) holds for every  $\ell \leq m$ , we prove it for  $m + 1$ . We only give the proof for the case  $m = 3m'$ ; the other cases follow in a similar way. In this case, by inductive assumptions and (4.4),

$$2(m + 1) + 1 = 3(2m' + 1) = 3(1 + f(i, t_{m'})) = 1 + f(2i + 1, 2t_{m'} + 1),$$

which implies that there is a solution of (4.6) for  $m + 1$ . Now, we prove the uniqueness. Let  $t \geq 2$  be a solution of (4.6) for  $m + 1$ . Then,

$$1 + f(i, t) = 2(m + 1) + 1 = 3(2m' + 1), \quad (4.7)$$

which implies  $f(i, t) \equiv 2 \pmod{3}$ . According to (4.4), this happens only if  $(i, t) \equiv (1, 1)$ . Write  $i = 2i' + 1$  and  $t = 2t' + 1$ . Then, by (4.4) and (4.7),

$$2m' + 1 = \frac{1 + f(i, t)}{3} = 1 + f(i' + 1, t').$$

By the inductive assumption, we know that  $t'$  is the unique solution of (4.6) for  $m'$ . So, the only solution of (4.6) for  $m + 1$  is  $2t_{m'} + 1$ .  $\square$

By Proposition 2, for every  $n \geq 2$ , we have

$$\mathcal{P}_{\mathbf{c}}^{(2)}(n) = M_{\mathbf{c}}(n - 2) + 2M_{\mathbf{c}}(n - 1) + 1 + \frac{1 + (-1)^{n+1}}{2}.$$

**4.2. Regularity of the  $k$ -abelian complexity of  $\mathbf{c}$ .** In this part, we prove the regularity of the  $k$ -abelian complexity of the Cantor sequence for every  $k \geq 3$ .

Let  $\mathcal{F}_{\mathbf{c}}$  denote the set of all factors of  $\mathbf{c}$ . For every  $u \in \mathcal{F}_{\mathbf{c}}$  and  $\ell \geq 1$ , we define

$$\text{Type}(\ell, u) := \{j = 0, 1, \dots, 3^\ell - 1 \mid u = c_{3^\ell n+j} \cdots c_{3^\ell n+j+|u|-1} \text{ for some } n \geq 0\}.$$

The elements in  $\text{Type}(\ell, u)$  are called *types* of  $u$  (with respect to  $\ell$ ). Clearly, for every  $\ell$  and  $u \in \mathcal{F}_{\mathbf{c}}$ ,  $\text{Card}(\text{Type}(\ell, u)) \geq 1$ .

Every type of  $u$  gives a decomposition of  $u$  in the following sense. For every  $j \in \text{Type}(\ell, u)$ , there is an integer  $n \geq 0$  such that

$$\begin{aligned} u &= (c_{3^\ell n+j} \cdots c_{3^\ell(n+1)-1}) (c_{3^\ell(n+1)} \cdots c_{3^\ell(n+h)-1}) (c_{3^\ell(n+h)} \cdots c_{3^\ell n+j+|u|-1}) \\ &= \text{suff}_{j_0}(\sigma^\ell(c_n)) \sigma^\ell(c_{n+1} \cdots c_{n+h-1}) \text{pref}_{j_1}(\sigma^\ell(c_{n+h})), \end{aligned} \quad (4.8)$$

where  $h = \lfloor \frac{|u|+j}{3^\ell} \rfloor$ ,  $j_0 = 3^\ell - j$  and  $j_1 = j + |u| - 3^\ell h$ . The following lemma shows that every non-zero factor of  $\mathbf{c}$ , which is long enough, occurs in a (relatively) fixed position, i.e., has only one type. By a *non-zero factor* we mean a factor that contains at least one letter '1'.

**Lemma 6.** *For every integer  $\ell \geq 1$  and every non-zero factor  $u \in \mathcal{F}_{\mathbf{c}}$  with  $|u| > 3^\ell$ ,*

$$\text{Card}(\text{Type}(\ell, u)) = 1.$$

*Proof.* We prove by induction on  $\ell$ . We first prove the result for  $\ell = 1$ . Now, we show that  $\text{Card}(\text{Type}(1, u)) = 1$  for  $u \in \mathcal{F}_{\mathbf{c}}(4)$  with  $|u|_1 > 0$ . We only verify the case  $u = 0001$  as an example; the rest can be verified in the same way. Suppose  $0001 = c_n c_{n+1} c_{n+2} c_{n+3}$ . Since  $c_{n+3} = 1$ , by (1.1), we have  $n \not\equiv 1 \pmod{3}$ . If  $n \equiv 2 \pmod{3}$ , then by (1.1),  $0 = c_{n+1} = c_{n+3} = 1$ , which is a contradiction. Thus,  $\text{Type}(1, 0001) = \{0\}$ .

For every non-zero factor  $u \in \mathcal{F}_{\mathbf{c}}$  with  $|u| > 4$ , let  $u = xvy$  where  $v$  is the first non-zero factor of length 4 of  $u$ . Since  $\text{Type}(1, v) + |x| \equiv \text{Type}(1, u) \pmod{3}$ , we have

$$\text{Card}(\text{Type}(1, u)) = 1.$$

Suppose the result holds for  $\ell$ . We prove it for  $\ell + 1$ . Let  $u \in \mathcal{F}_{\mathbf{c}}$  with  $|u| > 3^{\ell+1}$  and  $i_0 \in \text{Type}(\ell, u)$ . Then,

$$u = c_{3^\ell n + i_0} \cdots c_{3^\ell n + i_0 + |u| - 1}$$

for some  $n \geq 0$ . By (4.8),  $u$  uniquely determines  $i_0$ ,  $|u|$  and  $c_n c_{n+1} \cdots c_{n+h}$  where  $h = \lfloor \frac{|u|+i_0}{3^\ell} \rfloor$ . Since  $h \geq 3$ ,  $n \equiv i_1 \pmod{3}$  where  $i_1 \in \text{Type}(1, c_n \cdots c_{n+h})$ . Therefore,

$$3^\ell n + i_0 \equiv 3^\ell i_1 + i_0 \pmod{3^{\ell+1}}. \quad (4.9)$$

By the inductive assumptions,  $\text{Card}(\text{Type}(1, c_n \cdots c_{n+h})) = 1$  and  $\text{Card}(\text{Type}(\ell, u)) = 1$ . So, by (4.9), we have

$$\text{Card}(\text{Type}(\ell + 1, u)) = 1.$$

□

**Lemma 7.** *For every integer  $\ell \geq 1$  and every non-zero factor  $u \in \mathcal{F}_{\mathbf{c}}$  with  $3^\ell < |u| \leq 3^{\ell+1}$ ,*

$$1 \leq \text{Card}(\text{Type}(\ell + 1, u)) \leq 2.$$

*Proof.* Let  $u \in \mathcal{F}_{\mathbf{c}}$  with  $3^\ell < |u| \leq 3^{\ell+1}$  and  $i_0 \in \text{Type}(\ell, u)$ . Then,  $u = c_{3^\ell n + i_0} \cdots c_{3^\ell n + i_0 + |u| - 1}$  for some  $n \geq 0$ . By (4.8),  $u$  uniquely determines  $i_0$ ,  $|u|$  and  $c_n c_{n+1} \cdots c_{n+h} =: v$ , where  $h = \lfloor \frac{|u|+i_0}{3^\ell} \rfloor$ . Note that  $v$  is a non-zero factor. Write  $q(v) := \max\{j \mid 0^j \text{ is a prefix of } v\}$ . Then  $c_{n+q(v)} = 1$ , which implies  $n + q(v) \not\equiv 1 \pmod{3}$  by (1.1). So,

$$3^\ell n + i_0 \equiv -3^\ell q(v) + i_0 \text{ or } 3^\ell(2 - q(v)) + i_0 \pmod{3^{\ell+1}}. \quad (4.10)$$

The result follows from Lemma 6 and the above formula. □

In the rest of this section, let  $i$  be the integer satisfying

$$3^i + 1 < k \leq 3^{i+1} + 1.$$

To study the regularity of  $\{p_k(n, x, y)\}_{n \geq 1}$  for  $x, y \in \mathcal{F}_{\mathbf{c}}(k-1)$ , our idea is the following. We first give the upper bound of  $p_k(n, \cdot, \cdot)$  by using  $M_{\mathbf{c}}(\cdot)$ , which is a 3-regular sequence according to Corollary 1. Then, by constructing sufficiently many words which belong to different  $k$ -abelian equivalence classes, we show that the upper bound can be reached. Therefore, the regularity of  $\{p_k(n, x, y)\}_{n \geq 1}$  follows from the regularity of  $\{M_{\mathbf{c}}(n)\}_{n \geq 1}$ .

The following lemma contributes to the construction of words that belong to different  $k$ -abelian equivalence classes.

**Lemma 8.** *Let  $\alpha \in \{0, 1\}$ . For every  $\ell \geq 1$  and every  $h = 1, 2, \dots, M_{\mathbf{c}}(\ell)$ , there is a word  $W_h \in \mathcal{F}_{\mathbf{c}}(\ell + 3)$  such that  $|W_h|_1 = h$  and  $W_h = 00U_h\alpha$ , where  $U_h \in \mathcal{F}_{\mathbf{c}}(\ell)$ .*



*Proof.* For all  $j = 0, 1, \dots, \ell + 1$ , let

$$W_j = 0^{\ell+3-j} \text{pref}_j(\sigma^s(1)) \in \mathcal{F}_c(\ell + 3),$$

where  $s \in \mathbb{N}$  satisfying  $3^s > \ell + 1$ . Since  $|W_j|_1 \leq |W_{j+1}|_1 \leq |W_j|_1 + 1$  and  $|W_\ell|_1 = M_c(\ell)$ , we know that  $|W_j|_1$  changes from 0 to  $M_c(\ell)$  continuously while  $j$  takes values from 0 to  $\ell$ . So, for every  $h = 1, \dots, M_c(\ell)$ , there is a  $j_h (\leq \ell)$  such that  $|W_{j_h}|_1 = h$ . Moreover, we can require that the last letter of  $W_{j_h}$  is 0. Otherwise, 1 is the last letter of  $W_{j_h}$ . Then,  $W_{j_h+1}$  ends with 0 and  $|W_{j_h+1}|_1 = |W_{j_h}|_1$ .

There also is a  $j'_h$  such that  $|W_{j'_h}|_1 = h$ , of which the last letter is 1. Otherwise, 0 is the last letter of  $W_{j'_h}$ . Let  $m_h := \max\{q \mid 0^q \text{ is a suffix of } W_{j'_h}\}$ . Since  $|W_{j'_h}|_1 = h \geq 1$ , we always have  $m_h < j'_h$ . Then,  $W_{j'_h-m_h}$  ends with 1 and  $|W_{j'_h-m_h}|_1 = |W_{j'_h}|_1$ . If  $m_h > j'_h$ ,  $\square$

Now, we shall show the regularity of  $\{p_k(n, x, y)\}_{n \geq 1}$  for all  $x, y \in \mathcal{F}_c(k-1)$ .

**Lemma 9.**  $\{p_k(n, 0^{k-1}, 0^{k-1})\}_{n \geq 1}$  is a 3-regular sequence.

*Proof.* Without loss of generality, we can assume that  $n \geq 2 \cdot 3^{i+1} + 2k - 2$ , since changing finite terms of a sequence does not change its regularity. Noticing that  $3^i < k-1 \leq 3^{i+1}$ , the occurrence of each  $w \in \mathcal{W}_{n, 0^{k-1}, 0^{k-1}}$  in  $\mathbf{c}$  must be one of the four forms in Figure 1.

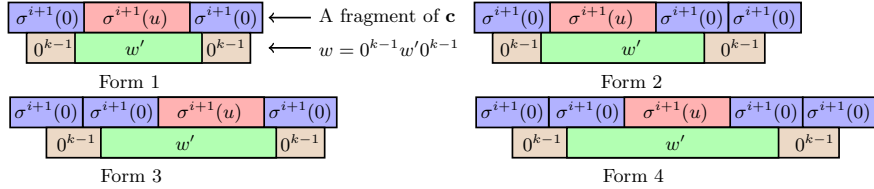


FIGURE 1.

In all the four forms, we have  $|w|_1 = 2^{i+1}|u|_1$  and  $|u| = \ell$  or  $\ell - 1$ , which implies

$$p_k(n, 0^{k-1}, 0^{k-1}) \leq M_c(\ell) + 1, \quad (4.11)$$

where  $\ell = \lfloor \frac{n-2k+2}{3^{i+1}} \rfloor$ . Next, we prove the inverse of (4.11). That is

$$p_k(n, 0^{k-1}, 0^{k-1}) \geq M_c(\ell) + 1. \quad (4.12)$$

Applying Lemma 8 for the above  $\ell$  and  $\alpha = 0$ , we have

$$W_h = 00U_h0 \in \mathcal{F}_c(\ell + 3) \text{ with } |W_h| = h$$

for all  $h = 1, 2, \dots, M_c(\ell)$ . Set  $t := n - 3^{i+1}\ell - k + 1$ . Then,  $k-1 \leq t < k-1 + 3^{i+1}$ . Therefore,

$$0^t \sigma^{i+1}(U_h) 0^{k-1} \in \mathcal{W}_{n, 0^{k-1}, 0^{k-1}} \text{ and } |0^t \sigma^{i+1}(U_h) 0^{k-1}|_1 = 2^{i+1}h$$

for every  $h = 1, \dots, M_c(\ell)$ . Noting also that  $0^n \in \mathcal{W}_{n, 0^{k-1}, 0^{k-1}}$ , the inequality (4.12) holds. The result then follows from (4.11), (4.12) and Corollary 1.  $\square$

For every non-zero factor  $v \in \mathcal{F}_c(k-1)$ , let  $z_v := \max\{p \mid 0^p \text{ is a suffix of } v\}$  and

$$\tilde{L}_v := \{q \bmod 3^{i+1} \mid c_{q-(k-2-z_v)} \cdots c_{q-1} c_q c_{q+1} \cdots c_{q+z_v} = v\},$$

where  $c_q$  is the last 1 in  $v$ . Then, it follows from Lemma 7 that  $1 \leq \text{Card}(\tilde{L}_v) \leq 2$ . Moreover, if  $\tilde{L}_v = \{q_1, q_2\}$ , where  $0 \leq q_1 < q_2 \leq 3^{i+1} - 1$ , then by (4.10), we have  $q_2 = q_1 + 2 \cdot 3^i$ .

For a word  $w = w_0 w_1 \cdots w_{n-1} \in \mathcal{A}^n$ , the reversal of  $w$  is defined to be  $\bar{w} = w_{n-1} \cdots w_1 w_0$ . When  $w = uv$ , we write  $wv^{-1} := u$  and  $u^{-1}w := v$  by convention.

**Lemma 10.** For all non-zero factors  $x, y \in \mathcal{F}_c(k-1)$ , two sequences  $\{p_k(n, 0^{k-1}, y)\}_{n \geq 1}$  and  $\{p_k(n, x, 0^{k-1})\}_{n \geq 1}$  are both 3-regular sequences.

*Proof.* For every  $x \in \mathcal{F}_c$ , its reversal  $\bar{x} \in \mathcal{F}_c$ , since  $x$  is a factor of  $\sigma^m(1)$  for some  $m \geq 1$  and  $\overline{\sigma^m(1)} = \sigma^m(1)$ . So,  $p_k(n, x, 0^{k-1}) = p_k(n, 0^{k-1}, \bar{x})$  for every  $n \geq 1$ . Thus, we only need to verify the regularity of  $\{p_k(n, 0^{k-1}, y)\}_{n \geq 1}$  for every non-zero factor  $y \in \mathcal{F}_c(k-1)$ .

Since changing finite terms of a sequence does not change its regularity, we can assume that  $n \geq 2 \cdot 3^{i+1} + 2k - 2$ . Recall that  $3^i < k - 1 \leq 3^{i+1}$ . Each occurrence of every  $w \in \mathcal{W}_{n, 0^{k-1}, y}$  in  $\mathbf{c}$  must be one of the six forms in Figure 2. In all the six forms, for every  $o_y \in \tilde{L}_y$ , we have

$$|w|_1 = 2^{i+1}|\tilde{u}|_1 - |\text{suff}_{3^{i+1}-o_y-1}(\sigma^{i+1}(1))|_1 := n_{o_y} \quad (4.13)$$

and  $|\tilde{u}| = \ell(o_y)$  or  $\ell(o_y) + 1$ , where

$$\ell(o_y) = \left\lfloor \frac{n - k - o_y - z_y}{3^{i+1}} \right\rfloor \quad \text{and} \quad \tilde{u} = \begin{cases} u01, & \text{if } w \text{ is of Form 5 or 6,} \\ u1, & \text{otherwise.} \end{cases}$$

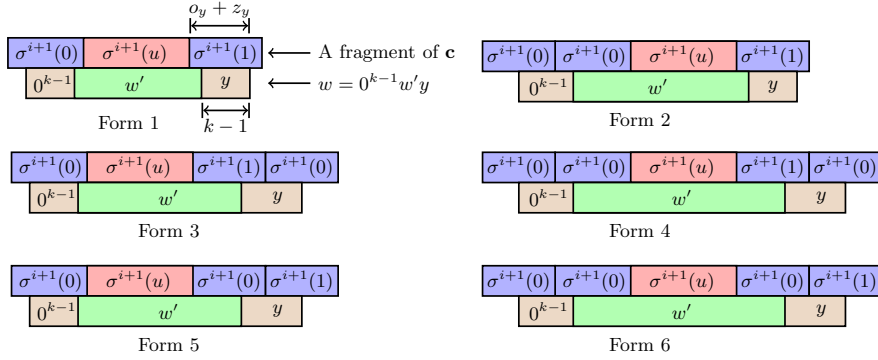


FIGURE 2.

When  $\text{Card}(\tilde{L}_y) = 1$ , write  $\tilde{L}_y = \{o_y\}$ . By (4.13), we have

$$p_k(n, 0^{k-1}, y) \leq M_c(\ell(o_y) + 1). \quad (4.14)$$

On the other hand, applying Lemma 8 for  $\ell(o_y)$  and  $\alpha = 1$ , we have

$$W_h = 00U_h1 \in \mathcal{F}_c(\ell + 3) \text{ with } |W_h|_1 = h$$

for all  $h = 1, \dots, M_c(\ell + 1)$ . Set  $t := n - 3^{i+1}\ell - o_y - z_y - 1$ ; so  $k - 1 \leq t < k - 1 + 3^{i+1}$ . Therefore,

$$V_{o_y} := 0^t \sigma^{i+1}(U_h) \text{pref}_{o_y+1}(\sigma^{i+1}(1)) 0^{z_y} \in \mathcal{W}_{n, 0^{k-1}, y}$$

and

$$|V_{o_y}|_1 = 2^{i+1}h - |\text{suff}_{3^{i+1}-o_y-1}(\sigma^{i+1}(1))|_1$$

for all  $h = 1, \dots, M_c(\ell + 1)$ . This implies that  $p_k(n, 0^{k-1}, y) \geq M_c(\ell(o_y) + 1)$ . The previous inequality, (4.14) and Corollary 1 give the result in the case  $\text{Card}(\tilde{L}_y) = 1$ .

Now suppose  $\text{Card}(\tilde{L}_y) = 2$  and set  $\tilde{L}_y = \{o_y, o'_y := o_y + 2 \cdot 3^i\}$  with  $0 \leq o_y \leq 3^i - 1$ . From (4.13), we know that  $n_{o'_y} \equiv n_{o_y} + 2^i \pmod{2^{i+1}}$ . Therefore,

$$p_k(n, 0^{k-1}, y) \leq M_c(\ell(o_y) + 1) + M_c(\ell(o'_y) + 1). \quad (4.15)$$

For every  $q \in \tilde{L}_y$ , applying Lemma 8 for  $\ell(q)$  and  $\alpha = 1$ , we have

$$W_{h,q} = 00U_{h,q}1 \in \mathcal{F}_c(\ell + 3) \text{ with } |W_{h,q}|_1 = h$$

for every  $h = 1, \dots, M_{\mathbf{c}}(\ell_1 + 1)$ . Set  $t(q) := n - 3^{i+1}\ell(q) - q - z_y - 1$ ; so  $k - 1 \leq t(q) < k - 1 + 3^{i+1}$ . Therefore, for every  $q \in \tilde{L}_y$ ,

$$V_q := 0^{t(q)}\sigma^{i+1}(U_{h,q})\text{pref}_{q+1}(\sigma^{i+1}(1))0^{z_y} \in \mathcal{W}_{n,0^{k-1},y}$$

and

$$|V_q|_1 = 2^{i+1}h - |\text{suff}_{3^{i+1}-q-1}(\sigma^{i+1}(1))|_1$$

for all  $h = 1, \dots, M_{\mathbf{c}}(\ell_1 + 1)$ . Since  $|V_{o_y}|_1 \equiv |V_{o'_y}|_1 - 2^i \pmod{2^{i+1}}$ ,  $V_{o_y}$  and  $V_{o'_y}$  belongs to different  $k$ -abelian equivalence classes. Therefore,

$$p_k(n, 0^{k-1}, y) \geq M_{\mathbf{c}}(\ell(o_y) + 1) + M_{\mathbf{c}}(\ell(o'_y) + 1). \quad (4.16)$$

Combining (4.15), (4.16) and Corollary 1, the result follows.  $\square$

**Lemma 11.** For two non-zero factors  $x, y \in \mathcal{F}_{\mathbf{c}}(k - 1)$ ,  $\{p_k(n, x, y)\}_{n \geq 1}$  is ultimately periodic.

*Proof.* Without loss of generality, we can assume that  $n \geq 2 \cdot 3^{i+1} + 2k - 2$  since changing finite terms of a sequence does not change its regularity. Noticing that  $3^i < k - 1 \leq 3^{i+1}$ , for every pair of factors  $x, y$  of length  $k - 1$ , the occurrence of each  $w \in \mathcal{W}_{n,x,y}$  in  $\mathbf{c}$  must be one of the nine forms in Figure 3.

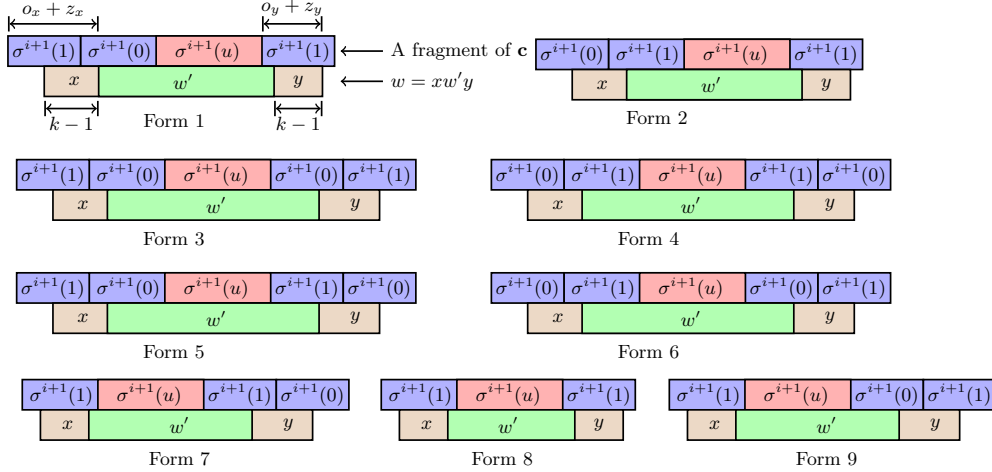


FIGURE 3.

For every fixed pair of  $o_x \in \tilde{L}_x$  and  $o_y \in \tilde{L}_y$ , in all the nine forms, we have

$$n = |w| = 3^{i+1}(|\tilde{u}| - 1) + \ell(o_x, o_y) \quad (4.17)$$

and

$$|w|_1 = 2^{i+1}|\tilde{u}|_1 - |\text{pref}_{o_x+z_x-k+2}(\sigma^{i+1}(1))|_1 - |\text{suff}_{3^{i+1}-o_y-1}(\sigma^{i+1}(1))|_1, \quad (4.18)$$

where  $\ell(o_x, o_y) := (k - 1 - o_x - z_x + o_y + z_y) < 2 \cdot 3^{i+1}$  and

$$\tilde{u} = \begin{cases} 10u1, & \text{if } w \text{ is of Form 1 or 5,} \\ 10u01, & \text{if } w \text{ is of Form 3,} \\ 1u01, & \text{if } w \text{ is of Form 6 or 9,} \\ 1u1, & \text{otherwise.} \end{cases} \quad (4.19)$$

Further, according to (4.5c),  $\tilde{u}$  in (4.19) must satisfy  $|\tilde{u}| \equiv 1 \pmod{2}$ . This fact and (4.17) yield that  $\mathcal{W}_{n,x,y} = \emptyset$  when  $n \not\equiv \ell(o_x, o_y) \pmod{2 \cdot 3^{i+1}}$ .

Now we deal with the case  $n \equiv \ell(o_x, o_y) \pmod{2 \cdot 3^{i+1}}$ . Note that by (4.5c), we have  $p(2j + 1, 1, 1) = 1$  for all  $j \geq 1$ . This fact and (4.18) imply that for all  $n = 2 \cdot 3^{i+1}j + \ell(o_x, o_y)$ ,

$$p_k(n, x, y) = \text{Card}(\{|w|_1 \mid w \in \mathcal{W}_{n,x,y}\}) = 1.$$

In conclusion, let  $\mathcal{I}_{x,y} = \{2 \cdot 3^{i+1}j + \ell(o_x, o_y) \mid j \geq 1, o_x \in \tilde{L}_x, o_y \in \tilde{L}_y\}$ . We have

$$p_k(n, x, y) = \begin{cases} 1, & \text{if } n \in \mathcal{I}_{x,y}, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore,  $\{p_k(n, x, y)\}_{n \geq 1}$  is ultimately periodic with a period  $2 \cdot 3^{i+1}$ . □

**Proposition 3.**  $\{\mathcal{P}_c^{(k)}(n)\}_{n \geq 1}$  is a 3-regular sequence for every  $k \geq 3$ .

*Proof.* It follows directly from Lemmas 9, 10 and 11 and (4.1). □

Theorem 2 follows from Propositions 1, 2 and 3.

## REFERENCES

- [1] J. P. Allouche and J. Shallit, *Automatic sequences: Theory, Applications, Generalizations*, Cambridge University Press 2003.
- [2] J. P. Allouche, J. Shallit, *The ring of  $k$ -regular sequences*, Theoret. Comput. Sci. 98 (2) (1992): 163-197.
- [3] L. Balková, K. Brinda and O. Turek, *Abelian complexity of infinite words associated with quadratic Parry numbers*, Theoret. Comput. Sci. 412 (45) (2011): 6252-6260.
- [4] M. Blake, and N. Rampersad, *The abelian complexity of the paperfolding word*, Discrete Math. 313(2013): 831-838.
- [5] J. Currie, N. Rampersad, *Recurrent words with constant Abelian complexity*, Adv. Appl. Math. 47(2011): 116-124.
- [6] S. Eilenberg, *Automata, Languages, and Machines*, vol. A, Academic Press, New York and London, 1974.
- [7] X. T. Lü, J. Chen, Y. J. Guo, and Z. X. Wen, *On the permutation complexity of the Cantor-like sequences*, Theoret. Comput. Sci. 616 (2016): 100-110.
- [8] J. Karhumäki, *Generalized Parikh mappings and homomorphisms*, Inform. and Control 47 (3) (1980): 155-165.
- [9] G. Richomme, K. Saari, L. Zamboni, *Abelian complexity of minimal subshifts*, J. Lond. Math. Soc. 83(2011): 79-95.
- [10] F. Blanchet-Sadri, N. Fox and N. Rampersad, *On the asymptotic abelian complexity of morphic words*, Adv. Appl. Math. 61 (2014): 46-84.
- [11] O. Turek, *Balances and abelian complexity of a certain class of infinite ternary words*, RAIRO Theor. Inform. Appl. 44 (3) (2010): 313-337.
- [12] F. Greinecker, *On the 2-abelian complexity of the Thue-Morse word*, Theoret. Comput. Sci. 593(2015): 88-105.
- [13] A. Parreau, M. Rigo, E. Rowland, E. Vandomme, *A New Approach to the 2-Regularity of the  $l$ -Abelian Complexity of 2-Automatic Sequences*, Electron. J. Combin. 22(2015): 1-27.
- [14] J. Karhumäki, A. Saarela, L.Q. Zamboni, *Variations of the Morse-Hedlund Theorem for  $k$ -Abelian Equivalence*, Developments in language theory, 203214, Lecture Notes in Comput. Sci., Vol. 8633, Springer, Cham, 2014.
- [15] J. Cassaigne, J. Karhumäki, A. Saarela, *On growth and fluctuation of  $k$ -abelian complexity*, *International Computer Science Symposium in Russia*, Springer International Publishing, 2015: 109-122.
- [16] J. Karhumäki, A. Saarela, L.Q. Zamboni, *On a generalization of Abelian equivalence and complexity of infinite words*, J. Combin. Theory Ser. A 120(2013): 2189-2206.
- [17] A. Parreau, M. Rigo, E. Vandomme, *A conjecture on the 2-abelian complexity of the Thue-Morse word*, <http://orbi.ulg.ac.be/handle/2268/162740>, 2014.

(J. Chen) COLLEGE OF SCIENCE, HUAZHONG AGRICULTURAL UNIVERSITY, WUHAN 430070, CHINA.  
*E-mail address:* `wind.golden@gmail.com`

(X.T. Lü) SCHOOL OF MATHEMATICS AND STATISTICS, HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY,  
WUHAN 430074, CHINA.  
*E-mail address:* `M201270021@hust.edu.cn`

(W. Wu) SCHOOL OF MATHEMATICS, SOUTH CHINA UNIVERSITY OF TECHNOLOGY, GUANGZHOU 510641, CHINA.  
*E-mail address:* `wuwen@scut.edu.cn`